806–813 Nucleic Acids Research, 1997, Vol. 25, No. 4

Role of proofreading and mismatch repair in maintaining the stability of nucleotide repeats in DNA

Bernard S. Strauss*, Daphna Sagher and Sonia Acharya

Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, IL 60637, USA

Received October 18, 1996; Revised and Accepted December 15, 1996

ABSTRACT

The role of the proofreading exonuclease in maintaining the stability of multiply repeated units in DNA was studied in *Escherichia coli*. Reversion of plasmids in which the β-galactosidase α complementing sequence was moved +2 out of frame by inserts containing $(CA)_{14}$, $(CA)_5$, $(CA)_2$ or $(TA)_6$ or +1 by creating a run of 8 C was compared in *mutS* and *mutSdnaQ* strains. Proofreading corrects at least half of the frameshift errors for all the plasmids and at least 99% of the errors in the $(CA)_2$ plasmid. The $(CA)_2$ plasmid reverts mostly by +1 frameshifts in the restriction sites flanking the insert. With the $(CA)_{14}$, $(TA)_6$, $(CA)_5$ and 8C plasmids, reversion is mainly by loss of a repeat unit. The data support the hypothesis that the *dnaQ* gene product recognizes frameshifts close to the DNA growing point. Frameshifts distal to the growing point are mainly corrected by mismatch repair. We speculate that mismatches in mononucleotide repeats are susceptible to proofreading because they can either migrate to a point where they are recognized by the exonuclease or, alternatively, because single nucleotide distortions are more readily detected than dinucleotides.

INTRODUCTION

Levinson and Gutman (1) showed that mismatch repair-deficient strains of *Escherichia coli* (*mutS* and *mutL*) produced increased numbers of frameshifts in (CA) repeat sequences (microsatellites) inserted into M13 bacteriophage DNA. Strand *et al*. (2), working with yeast, confirmed the observation that instability was associated with a deficiency in mismatch repair. Yeast mutants deficient in mismatch repair deleted a (CA) repeat unit several hundred fold more frequently than the wild-type. The frequency of (CA) deletion was either unchanged or increased a modest 10-fold in yeast strains deficient in the proofreading exonuclease of either of the two DNA polymerases as compared with the wild-type. They suggested that either the proofreading exonuclease did not detect loops or bubbles that form away from the growing point or that heterologies of >1 bp were not corrected (2). Although both yeast DNA polymerase mutants are efficient mutators [in contrast to *E.coli* pol I mutants; (3)], the effect of double mutants, deficient in both replicative exonucleases, on frameshift mutation was not reported.

We previously demonstrated a role for proofreading exonuclease activity in UV-induced frameshift mutagenesis at repeated sequences *in vitro* (4). There is also *in vitro* evidence that proofreading exonuclease is important in the surveillance of spontaneous frameshift mutations (5). Since pol III is the major *E.coli* replication polymerase, the use of mutants deficient in the exonuclease subunit of this enzyme offered an interesting additional test of the possible role of proofreading in microsatellite instability. The experiments reported here with *dnaQ* mutants show that, as in yeast (2), proofreading has only a minor role in the surveillance of frameshifts in long repeated sequences. The role of proofreading depends on the size and composition of the repeats. Since the exonuclease acts at DNA growing points, whereas mismatch repair occurs at a distance from the replication fork, these experiments imply that the events in frameshift mutation at highly repeated sequences are not limited to the DNA growing point.

MATERIALS AND METHODS

Except where otherwise indicated, the methodology and the media used are as described by Miller (6).

Strain construction

The selectable marker *met D* is located close to the replicative polymerase (*dnaE*) and proofreading (*dnaQ*) loci. We constructed a strain which carried *metD*$^-$*metB*$^-$ along with a deletion in the *lac* region. This strain, termed BS40 [*metD*$^-$*metB*$^-$ *ara*$^+$ *pro*$^-$ Δ(*lac*) *Str*^{r}] was the progeny of a cross of S90C [F$^-$ *Str*^{r} *metB*$^+$ *ara*$^-$ *metD*$^+$ Δ(*pro lac*)] × Hfr CD4 (*Str*^{s} *metB*$^-$ *ara*$^+$ *met D*$^-$ *proA*$^-$) selected on arabinose + streptomycin-containing plates with subsequent screening for a *pro*$^-$ colony unable to utilize D-methionine. This strain must carry a deletion in the *lac* region. We have not determined whether the deletion comes from the CD4 or S90C parent. The newly isolated mutators (see below) were transferred into BS40 by P1 transduction, selecting for *metD*$^+$ and replica plating onto rifampicin plates to identify mutators. The fertility factor of XL-1 blue (F′::Tn*10* *proA*$^+$*B*$^+$*lacI*^{q}*lacZ*Δ*M15*) was transferred into BS40 and its derivatives to allow for M13 growth. The *mutS* (*mutS215*::Tn*10*) and *mutL* (*mutL218*::Tn*10*) strains prepared by E.Siegel were obtained from the *E.coli* Genetic Stock Center in a *thy*$^-$ background. Since these strains were susceptible to thymineless death, we transduced *mutS* and *mutL* into strain S90C by selection with streptomycin and tetracycline. S90C

*To whom correspondence should be addressed. Tel: +1 773 702 1628; Fax: +1 773 702 3172; Email: bs19@midway.uchicago.edu

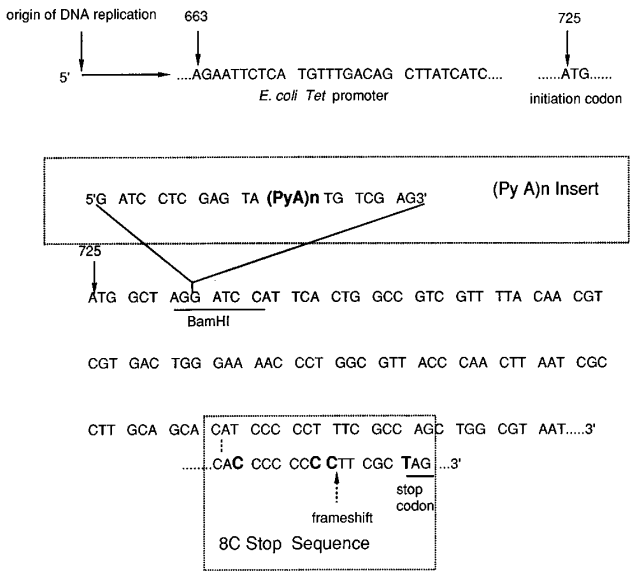

Figure 1. Sequence changes in plasmid and phage *lacZ* inserts. Numbering represents the distance from the end of the RNA primer during replication from the pBR322 origin (25).

mutH was obtained from Dr J.Miller. A list of the strains used in this investigation is available in the on-line edition of this Journal.

Plasmid construction

Repeat sequences (Fig. 1) were introduced into the β-galactosidase gene in both M13mp2 and in a plasmid, putting the gene out of frame and yielding colorless plaques or colonies when grown in appropriate strains. Reversions were detected as blue colonies or plaques. The plasmid used was p205-GTI (7) This 8.6 kb plasmid has an SV40 origin and a G418 resistance marker for selection in mammalian cells, a pBR322 origin for replication in bacteria and an Amp resistance marker. The LacZ α peptide differs from the wild-type in codons 2–5 and is preceded by the *E.coli tet* promoter. The sequence at codons 3 and 4 was changed to create a *Bam*HI site for insertion of the different oligomers containing the PyA repeats. These oligomers were based on the sequence of pSH31 (8), with *Bam*HI sites at the ends. Modifications of the *lacZ* sequence (creation of the *Bam*HI site or the 8C stop sequence) were by *in vitro* mutagenesis in uracil-containing M13 (9). The 8C stop sequence (Fig. 1) was prepared using the oligonucleotide CACCCCCCCCTTCGCTAG to replace the wild-type CATCCCCCTTTCGCCAG at codons 30–35 of the modified *lacZ*. M13 RF was then prepared, digested with *Bcl*I and *Bgl*II, separated on an agarose gel and the modified sequence isolated from the gel and ligated into a similarly digested plasmid. The M13 used for these manipulations was derived from M13mp2 by replacing the *Ava*II–*Pvu*I fragment (positions 5914–6351) with the *Bgl*II–*Bcl*I fragment from plasmid p205-GTI (modified *lacZ* and *tet* promoter). All experiments described in this study were done with the same basic sequence in both the phage and the plasmid. We designate the M13 phage containing the $(CA)_{14}$ insert M13is$(CA)_{14}$. The modified plasmids containing the different repeats are called $(CA)_{14}$, $(CA)_5$, $(CA)_2$, $(TA)_6$ [since a $(TA)_5$ is placed next to a TA in the basic insert] and 8C stop.

Mutagenesis of P1 phage

P1 phage were grown on strain CD4 transduced to *metD*$^+$. A phage suspension (4×10^{13} p.f.u./ml) was mutagenized with 1.0 M hydroxylamine solution (10). Phage survival was 2 and 6% in different runs measured against a mock-treated control. The mutagenized phage preparation was used for transduction (6). Transductants were isolated on minimal agar plates containing D-methionine (10 μg/ml), 2 mM phenyl-β-D-galactoside (P-gal), 150 μM 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), streptomycin (125 μg/ml) and citrate (20 mM, reduced to 5 mM in later experiments; 6). This selective medium is designed to identify mutants which form blue papillae (11).

Sequencing *dnaQ* mutants

Three overlapping fragments of cellular DNA were amplified with primer pairs based on sequences generated with the MacVector program. Numbering starts at the first nucleotide of the GenBank *dnaQ* sequence (Locus ECORNHQ, accession no. K00985M3020). The primer pairs were: F10 (283–304) 5′-TGATACCCTGGCGGACATACTG-3′, B9 (842–820) 5′-CATGAACTCATCGGCTACTTCGG-3′; F12 (577–600) 5′-CCGCTATGAGCACTGCAAT TACAC-3′, B11 (1103–1080) 5′-AACTTCCGCAAGGATCTGGGCATC-3′; F23 (893–916) 5′-CGGCTTTATGGACTACGAGTTTTC-3′, B32 (1529–1508) 5′-AGTGAATAGTGGCGGAACGGAC-3′. PCR products were sequenced using the Promega fmol TM*1 Sequencing System with the amplification primers and the additional following internal primers: B3 (626–603) 5′-GGTTTCGGTATCGAGAACGATCTG-3′; B15 (1168–1145) 5′-TGCTGTTGTGTCTCTCCTTCCATC-3′; F10b (367–386) 5′-CCATCAACTCCATACGGGTTG-3′; F17b (678–696) 5′-ATTGGTGCCGTTGAAGTGG-3′; F22b (1011–1029) 5′-AGCCTCGATGCGTTATGTG-3′; F27b (1219–1236) 5′-GCGTTGTTTTTGCGACAG-3′. Cycle sequencing reactions were done with ^{32}P-end-labeled primers for 30 cycles or with an ABI prism automatic sequencer at The University of Chicago Cancer Research Center.

Reversion analysis

Single plaques of M13mp2 phage containing the appropriate constructs were picked into 1.5 ml LB medium and grown at 37°C overnight. The supernatants were assayed on strain JM103. Plasmids were introduced into host cells by $CaCl_2$-mediated transformation and the cultures were grown overnight or, for *mutSdnaQ* mutants, until visible growth was obtained. Plasmid preparations were introduced into strain JS5 by electroporation with a BioRad Gene Pulser with Pulse Controller set at 2.5 kV and with 25 μFD capacitance at 400 Ω using 0.2 cm cuvettes. The medium for the detection of reversion to complementation of the β-galactosidase genes is LB + 'A salts' (6) supplemented with 100 mM IPTG, 150 μM X-gal and ampicillin (100 μg/ml). Blue colonies were scored after at least 18 h incubation at 37°C.

Revertants were colony purified. $(CA)_2$ revertants segregated colorless colonies even after several subcultures and many of the plasmid preparations from this construct were eventually sequenced as mixtures. Plasmid DNA was isolated from a 5 ml overnight culture of purified colonies. Sequences were obtained either by cycle sequencing with the ABI Prism 377A Sequencer (courtesy of the UCCRF-DSF) or manually with the USB Sequenase Version 2.0 Kit for dideoxy Sanger sequencing with [α-^{35}S]dATP (>1000 Ci/mmol; Amersham) and dITP mixes. The primer for the

automatic cycle sequencing reactions was 5′-CTGCGTGCAATCCATCTTGT-3′, which primes 135 bp away from the insert. The primer for the Sanger sequencing was 5′-TGGGTAACGCCAGGGTT-3′, which primes 35 bp from the insert.

Quantitative PCR

Quantitative PCR was carried out on total DNA using ^{32}P-end-labeled primers designed to amplify either a 265 bp target in the *dnaQ* gene (chromosomal marker, primers 5′-CATGAACTCATCGGCTACTTCGG-3′ and 5′-AACTTCCGCAAGGATCTGGGCATC-3′) or a 460 bp target in the plasmid [p205 $(CA)_5$] extending from a section of the kanamycin resistance gene on one side (5′-CCTGCGTGCAATCCATCTTGTTC-3′) to the eukaryotic thymidine kinase on the other (5′-TCCACTTCGCATATTAAGGTGACG-3′). Bacterial cultures (1.5 ml) were spun down, washed in water and then resuspended in 50 µl water, mixed with 150 µl 5% Chelex (Chelex 100 Resin, 100–200 Mesh, sodium form; BioRad) in water. The suspension was boiled for 10 min, chilled on ice, centrifuged for 15 min at top speed in a microfuge and the supernatant diluted in water. Diluted supernatant (5µl) containing 6–60 pg DNA (for plasmid amplification) or 60–150 pg (for chromosomal amplification) was subjected to PCR in the presence of 2 pmol ^{32}P-end-labeled forward primer and 4 pmol backward primer using *Taq* polymerase. Following 5 min at 94°C, amplification was for 20 cycles of 1 min at 57°C (annealing), 1 min at 72°C (extension) and 1 min at 94°C, with a final 10 min extension. Samples were separated on a 7.2% acrylamide gel and the amplified bands quantified with a phosphorimager (STORM 860; Molecular Dynamics). The DNA amounts used were such that a linear increase in concentration gave a linear increase in the intensity of the signal. For each target, only one primer was end-labeled. The primers for the different targets were labeled with the same [^{32}P]ATP mix so a direct comparison of number of molecules could be made from the relative radioactivity of the target bands.

Mutation rate calculations and statistical treatment of the data

Mutation rates (µ) were calculated from the median frequency by reiteratively solving the equation: $\mu = 0.4343\, f/\log(n\mu)$, where f is the median mutation frequency and n is the population size (12). The gene being studied is located on a plasmid and population size should properly be related to plasmid copy number. Although we do not have absolute values, we show below that the *mutS* and *mutSdnaQ* strains used contain similar plasmid copy numbers. Since the calculated mutation rate is relatively insensitive to changes in the value of n, we used viable count rather than estimated plasmid number as a measure of population size. There is only a 1.3-fold difference in calculated mutation rate for a 100-fold difference in population size (Table 3). The mutation rate is a calculated value and we use the experimentally determined median mutation frequency for statistical analysis. For this analysis we employed the Wilcoxon rank sum test, a non-parametric statistical method which does not depend on a normal distribution of the data. Comparison of the ratio of *mutSdnaQ/mutS* frequencies among the different plasmids was then performed using analysis of variance after a log transformation across the different groups (plasmids) to stabilize the variance (13).

RESULTS

dnaQ mutants

We took advantage of the proximity of *dnaQ* and *dnaE* to the *metD* locus to isolate mutants by the localized mutagenesis technique (10). Mutagenized P1 phage were used to transduce a *metD*$^-$*metB*$^-$ Δ(*proAB lac*) *Str*^{r} strain carrying the CC101 F′ factor which reverts from *lac*$^-$ to *lac*$^+$ by an A:T→C:G transversion (14). *Met*$^+$ transductants were selected and putative mutators were recognized among the transductants by the presence of numerous blue papillae against a colorless background. Approximately 30 putative mutators (*dnaQ-A*, *-B*,...), all from separate plates and from different transductions, were isolated. The isolates were purified, checked for a mutator phenotype on rifampicin plates and the mutation transferred by P1 transduction to strain BS40, selecting for *metD*$^+$. Transductants were first transferred onto LB plates and then replica plated to rifampicin plates for classification of mutator activity.

Table 1. Sequence changes observed in *dnaQ* mutants

Mutant	Rif^r $f \times 10^4$	Change	GenBank nucleotide no.	Exonuclease domain (14)	Amino acid Original	Mutant	No.	Codon
dnaQ-A	1.3	C→T	1252	III	Ala	Val	223	GCT→GTT
dnaQ-B	0.55	G→A	621	I	Glu	Lys	14	GAA→AAA
dnaQ-C	0.93	G→A	1080	int II–III	Asp	Asn	167	GAT→AAT
dnaQ-E	0.11	G→A	1080	int II–III	Asp	Asn	167	GAT→AAT
dnaQ-K	0.19	G→A	1080	int II–III	Asp	Asn	167	GAT→AAT
dnaQ-V	0.83	G→A	1080	int II–III	Asp	Asn	167	GAT→AAT
dnaQ-AB	0.15	G→A	1080	int II–III	Asp	Asn	167	GAT→AAT
dnaQ-F	2.3	G→A	888	II	Asp	Asn	103	GAT→AAT
dnaQ-H	2.5	G→A	888	II	Asp	Asn	103	GAT→AAT
dnaQ-U	2.4	G→A	888	II	Asp	Asn	103	GAT→AAT
dnaQ-G	0.92	C→T	1065	int II–III	His	Tyr	162	CAC→TAC
dnaQ-L	0.37	G→A	615	I	Asp	Asn	12	GAT→AAT
dnaQ-N	0.027	C→T	1072	int II–III	Ala	Val	164	GCA→GTA
dnaQ-O	1.3	C→T	1072	int II–III	Ala	Val	164	GCA→GTA
dnaQ-P	1.5	G→A	631	I	Gly	Asp	17	GGT→GAT

f, frequency of rifampicin-resistant mutants. The value for the *dnaQ*$^+$ strain (BS40) is 5×10^{-9}.

Total DNA was extracted from the different isolates and *dnaQ* gene segments were amplified and sequenced in both directions as described in Materials and Methods. Out of 15 isolates studied there were eight different *dnaQ* alleles (Table 1). *DnaQ-N* and *dnaQ-O* are identical to the mutation described as *mutD51* (15). Five alleles are within the ExoI, ExoII and ExoIII regions defined as conserved in 3′→5′ exonucleases (16). Three, including mut*D51*, are clustered midway between the ExoII and ExoIII domains. We have not found a secondary mutation elsewhere in the genome to account for the lower mutability of *dnaQ-N* (Table 1). Most of the experiments described below utilize *dnaQ-G* (codon 162) or *dnaQ-E* (codon 167). Plasmids carrying either the wild-type allele of *dnaQ* (pMM5;17), *dnaE* [pOPPE (18,19) or a control insert $p(TA)_6$ (this paper)] were introduced into *dnaQ-E* and *dnaQ-G* to test complementation by lowered rifampicin or nalidixic acid resistance. Only the *dnaQ*$^+$ plasmid showed complementation of *dnaQ-E* and *dnaQ-G*, consistently lowering mutation frequency in the *dnaQ* mutants by about two orders of magnitude (data not shown). The mutants used in these studies had been transferred twice by P1 transduction and *metD* selection since their original isolation by P1 mutagenesis. It is therefore unlikely that they carry any additional mutator mutations outside the *dnaQ* locus (20).

Experiments with bacteriophage M13

We determined the effect of both *dnaQ* and mismatch repair mutations on reversion of a $(CA)_{14}$ sequence incorporated into the *lacZ* region of a modified M13mp2. We found a frequency of reversion to lac$^+$ of ~1% when the $(CA)_{14}$-containing phage was grown in wild-type bacteria (Table 2). This wild-type frequency was higher than previously reported (1) for a longer repeat sequence [$(CA)_{21}$]. When the $(CA)_{14}$-containing phage were grown on *mutH*, *mutL* or *mutS* strains an increased frequency of reversion, although not as large as previously reported for the longer sequence (1), was observed. *DnaQ* mutants also displayed a higher than wild-type reversion frequency (Table 2).

Experiments with plasmids

In order to obtain an experimental system in which the magnitude of the effect of mismatch repair deficiency was closer to that observed in yeast (2), we prepared a series of repeated sequences contained in plasmid molecules. Mutation rates were compared in wild-type, *mutS* and *mutSdnaQ* strains in order to reduce the uncertainty in interpretation caused by the functional mismatch repair deficiency of *E.coli dnaQ* mutants (21,22). *DnaQ* and *mutSdnaQ* mutants were prepared from BS40 by successive P1 transductions, transformed with the particular plasmid and subcultured without purification of the transformants by restreaking. Whole colonies of the *mutSdnaQ* double mutants were lifted from the plates and inoculated into 10 ml LB + ampicillin to minimize the accumulation of suppressor mutations in these mutable strains (20). Cultures of the double mutants are always heterogeneous and we avoided picking the larger colonies for transformation. It took between 24 and 36 h to obtain dense cultures of transformed *mutSdnaQ* double mutants for plasmid preparation. About 10–20% of the picked colonies failed to grow. Plasmid preparations from all strains were assayed for reversion by electroporation into strain JS5.

Table 2. Lac$^+$ frequency after growth of M13is$(CA)_{14}$ in mutator strains

Strain	Rif$^r \times 10^7$	Blue plaques (%)	Median
S90C	0.17	1.2	1.2
		1.3	
		0.84	
BS40*metD*$^+$/XL(a)	0.24	1.3	1.3
		1.0	
		4.9	
BS40*metD*$^+$/XL(b)	0.06	0.71	1.0
		1.0	
		1.0	
S90*mutH*/XL	26	3.3	3.8
		3.8	
		4.2	
S090*mutL*/XL	26	5.3	5.3
		4.9	
		12.	
S90*mutS*/XL	20	5.2	4.7
		4.7	
		2.6	
BS40*dnaQ*-E/XL	2000	3.2	3.5
		3.5	
		4.2	
BS40*dnaQ-G*/XL	2500	3.0	3.8
		3.8	
		5.2	
BS40d*naQ-N*/XL	720	2.7	2.7
		2.7	
		4.0	

Single plaques were picked and grown in 1.5 ml LB medium in triplicate. Cultures were incubated overnight. The supernatants were collected and assayed on strain JM103.
(a) and (b) represent independent isolates of BS40*metD*$^+$.
Rif$^r \times 10^7$, frequency of rifampicin-resistant bacteria in the different host strains.

DNA replication in plasmids with origins derived from pBR322 (ColE1) is unidirectional and starts by the formation of a long RNA transcript which is then replicated for 200–400 bases by *E.coli* pol I, after which replication on the leading strand switches to pol III (23). Lagging strand replication is always by pol III. The instability of long trinucleotide repeats in *E.coli* is greatest when inserts are close to the origin; CTG repeats inserted 0.2–0.45 kb downstream of the origin were reported to be 'much less' stable than when inserted 1.5 kb away (24). Our inserts were all downstream of the initiation codon for the α peptide, which starts 725 bases from the first deoxynucleotide added in DNA replication (25).

We constructed plasmids with 14, five or two (CA) repeats, an oligonucleotide with five (TA) repeats, which when inserted into our vector gave a sequence with six (TA) repeats, and a sequence with eight C residues followed by a stop codon (Fig. 1). The dinucleotide repeats shifted the α peptide out of frame by +2/–1 nt. The shift produced by the mononucleotide repeat (8C stop) is +1. Mutation rates were calculated by the method of Drake (12) using the median reversion frequency in a series of independent cultures to minimize the effect of outliers. The rate of reversion to a lac$^+$ phenotype decreases rapidly as the number of repeated elements in the insert is diminished (Table 3). The rate is very much higher for plasmid molecules propagated in a *mutS* strain.

Table 3. Mutation rates and frequencies for plasmid inserts

Strain	Plasmid insert				
	$(CA)_{14}$	(8C stop)	$(TA)_6$	$(CA)_5$	$(CA)_2$
Wild-type BS40					
Mutation rate	4.9×10^{-5}	7.0×10^{-7}	4.2×10^{-7}	9.5×10^{-7}	8.6×10^{-8}
Median frequency	5.9×10^{-4}	5.2×10^{-6}	2.8×10^{-6}	7.6×10^{-6}	4.7×10^{-7} [a]
No. of samples	12	9	10	15	17
mutS					
Mutation rate	5.4×10^{-3}	8.3×10^{-4}	2.5×10^{-4}	1.7×10^{-4}	4.0×10^{-8}
Median frequency	8.8×10^{-2}	1.2×10^{-2}	2.8×10^{-3}	2.0×10^{-3}	1.5×10^{-7} [a]
No. of samples	11	14	13	22	17
mutSdnaQ-G					
Mutation rate	9.0×10^{-3}	4.3×10^{-3}	1.1×10^{-3}	3.8×10^{-4}	3.5×10^{-6}
Median frequency	13.4×10^{-2}	5.8×10^{-2}	1.2×10^{-2}	4.4×10^{-3}	1.8×10^{-5}
No. of samples	12	17	12	19	17

*15 of 17 wild-type and 13 of 17 *mutS* cultures had 0 revertants, giving a median frequency of 0. The value given is the total number of revertants in all 17 replicates divided by the total number of colonies screened.

The introduction of a *dnaQ* mutation into the *mutS* strains makes for a dramatic increase in the reversion of the $(CA)_2$ plasmid and for a 4- to 5-fold increase in the reversion rate of the $(TA)_6$ and 8C stop plasmids. The increase in $(CA)_5$ is smaller and the effect of *dnaQ* on reversion of the $(CA)_{14}$ plasmid appears smaller still. The measured mutation frequencies rather than the calculated mutation rates have been used for the statistical analysis of differences between the reversion of the constructs in the different strains. Specifically, comparison of *mutS* with *mutSdnaQ* groups using the Wilcoxon (non-parametric) rank sum test yielded $P < 0.0001$ for the $(CA)_2$, $(CA)_5$, $(TA)_6$ and 8C stop plasmids and $P < 0.0017$ for the $(CA)_{14}$ plasmid. We conclude that there is an effect of proofreading on the stability of all the repeat tracts. The estimated ratios (actually the ratio of the geometric means) of the *mutSdnaQ* and *mutS* frequencies were 2.15, 3.29, 3.92 and 5.39 for the $(CA)_{14}$, $(CA)_5$, $(TA)_6$ and 8C stop plasmids respectively. Pairwise comparisons detected a significant difference between the 8C stop and $(CA)_{14}$ ratios ($P = 0.008$), but none of the other pairwise comparisons reached statistical significance. A test of the hypothesis that the *mutSdnaQ*/*mutS* frequency ratio is different for the 8C stop and $(TA)_6$ plasmids combined as compared with the $(CA)_{14}$ and $(CA)_5$ plasmids (i.e. the average of the two former versus the average of the two latter) yielded $P = 0.020$. Due to the multiple comparisons involved in the statistical analysis we think this hypothesis needs to be confirmed by additional, independent experiments. Our data with the *dnaQ-E* strain (data not given) are similar but not extensive enough to permit statistical analysis.

Revertant analysis

As expected for $(CA)_{14}$, $(CA)_5$ and $(TA)_6$, the major change is loss of a single dinucleotide unit (Table 4). No +1 revertants were observed. The 8C stop plasmid was constructed so that revertants could be obtained by loss of a single nucleotide and the 18 independent revertants sequenced had –1 changes. A more complex pattern was observed among the revertants of the $(CA)_2$ plasmid (Fig. 2). One revertant, found after propagation of the plasmid in a *mutSdnaQ-E* strain, had addition of an A in the CACA repeat region. The majority of the revertants were +1 duplications of single nucleotides, observed for the most part in the two *Bam*HI restriction sites flanking the inserted sequence. Figure 2, indicating the location of these mutations, has been drawn to suggest a possible stem–loop in the vicinity of the restriction sites.

Table 4. Sequence changes in revertants

Plasmid insert	Change ΔCA	+1 *n*	–1 *n*	Other	Total
$[CA]_{14}$	52*			2 (–4CA)	
				2 (+5CA)	
				1 (+CA, no other change detected)	
				1 (–5CA, –6 other, +2)	
				*Includes a G→A	58
$[TA]_6$	7 (ΔTA)				7
$[CA]_5$	48*			*Includes 2 C→T	50
				2 (Δ complete insert)	
8C stop			18		18
$[CA]_2$		54		1 no insert	
				1 duplication 723–744	56

Plasmid copy number and mutation rate

Calculation of mutation rates requires knowledge of the number of replications, which is usually estimated by a count of viable cells. However, there is evidence (see below) that cultures of the double mutant contain a significant proportion of non-viable cells. It seemed possible that the dead *dnaQmutS* cells accumulated intact plasmid copies or that the viable double mutant cells had a higher plasmid copy number. The relative number of plasmid to

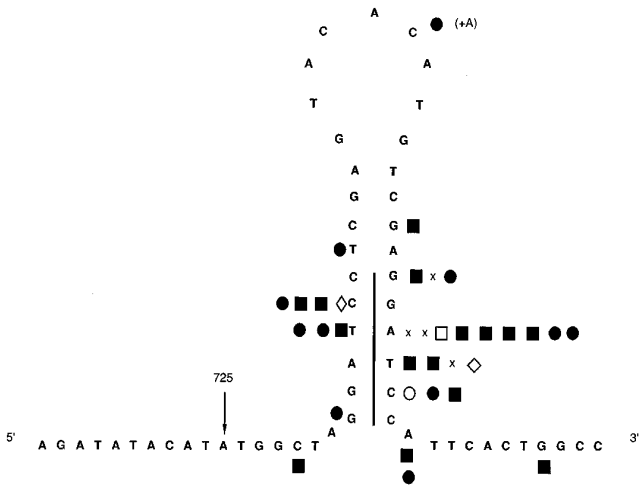

Figure 2. Location of $(CA)_2$ mutations. Each independent $(CA)_2$ revertant sequence is indicated on this portion of the p205$(CA)_2$ sequence. The heavy vertical line identifies the two *Bam*HI sites in the insert. Possible sibs have been eliminated. All symbols indicate duplication of the indicated nucleotide except when otherwise specified. Revertants were isolated from wild-type (cross), *mutS* (diamonds), *dnaQ-G* (open squares), *dnaQ-E* (open circles), *mutSdnaQ-G* (closed squares) and *mutSdnaQ-E* (closed circles) strains.

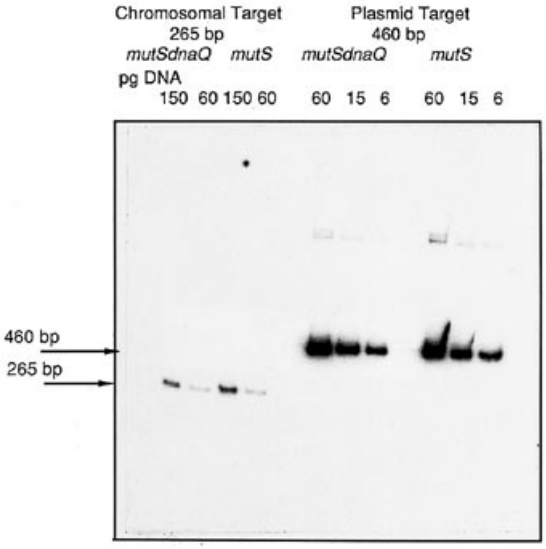

Figure 3. PCR amplification of plasmid and chromosome sequences. Phosphorimager scan of the amplification products of DNA from strains BS40*mutS* and BS40mut*SdnaQ-G*, both containing p205$(CA)_5$. No 460 bp band was detected when cells without plasmid were subjected to amplification with the same primers.

chromosome copies was determined by adapting a PCR method developed by S.Benson (personal communication). BS40*MutS* and BS40*mutSdnaQ* containing the p205$(CA)_5$ plasmid were grown for 12 and 40 h respectively. Total DNA was extracted and plasmid and chromosomal sequences in the same samples were amplified and the products quantified after electrophoresis (Fig. 3).

There are ~60 plasmid copies/chromosome target in *mutS* for each 100 plasmid copies/chromosome in the *mutSdnaQ* double mutant. This comparison of the two strains is valid irrespective of the efficiency of amplification of the two targets, since each primer pair should have the same efficiency with DNA preparations from different strains. The 100-fold difference in the viable count (2.4×10^9/ml for *mutS* versus 2.3×10^7/ml for *mutSdnaQ*) compared with only a 6-fold difference in the DNA content of the lysates (45 µg in the *mutS* sample versus 7.8 µg in *mutSdnaQ*) indicates that the *mutSdnaQ* culture contains large numbers of dead cells, as would be expected for organisms making so many errors (26). Using the amount of DNA recovered as a measure of the total cell number, the calculated number of plasmid divisions in the *mutS* cultures is actually somewhat greater than in the double mutant (45×60 compared with 100×7.8). There is no evidence that the higher mutation rate in *mutSdnaQ* strains results from a greater number of plasmid copies.

DISCUSSION

Repeated sequences in plasmids are subject to efficient mismatch repair

The frequency of reversion of the $(CA)_{14}$ repeat in phage grown in mismatch repair-deficient strains was 3- to 5-fold greater than when grown in wild-type hosts (Table 2). Levinson and Gutman (1) reported 16-fold increases for $(CA)_{21}$ repeats in their phage. This low ratio in the bacteriophage as compared with yeast (2) results from the relatively high control values of ~1% in the wild-type host (Table 2). One possible explanation is that the large number of slippage events in many phage DNA molecules saturates the mismatch repair system so that only part of the frameshifts can be corrected (21). A second possibility is that during single-stranded replication DNA is encoated in protein before the mismatch repair system has had time to act.

The mutation rate for $(CA)_{14}$ plasmids in a *mutS* strain is 110 times higher than in the wild-type. The ratio of reversion in *mutS* compared with the wild-type is >1000 for the 8C stop plasmid. These values compare with ratios of 100 (for chromosomal) and 500–700 for plasmid repeat sequences in yeast (2) and indicate that frameshifts in plasmid DNA are very efficiently monitored by the *E.coli* mismatch repair system.

The wild-type *dnaQ* gene product detects frameshifts

The *dnaQ-G* and *dnaQ-E* mutants were isolated on the basis of an inability to detect transversions. It is likely that frameshifts occur as a result of failed interactions between polymerase and nucleic acid substrate at one domain (27), whereas base substitutions arise as a result of errors at another domain, most probably the growing point region. It might well be that different exonuclease changes have a differential effect on the capacity to proofread base substitution and frameshift errors. Our data show that the *dnaQ-G* mutant used in this work (and the *dnaQ-E* mutation, mapped only five codons away, for which we have less extensive data) plays a major role in the recognition of some frameshifts.

The total number of errors made by the polymerase can be obtained from the mutation rate in the *mutSdnaQ* strain, in which both proofreading and mismatch repair are absent. The value for mutation rate obtained from the *mutS* strain gives the number of errors made in the absence of mismatch repair but with efficient proofreading. The relative contribution of proofreading can be determined by comparison of mutation rates in the *mutS* and *mutSdnaQ* strains. A plot of the rate of polymerase errors for the

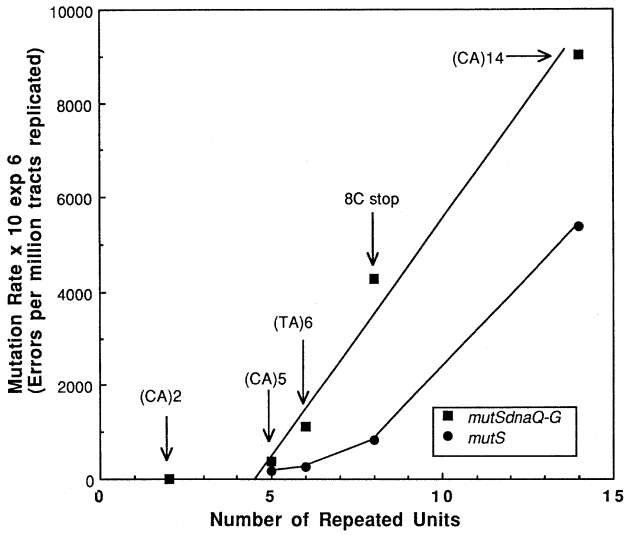

Figure 4. Errors per 10^6 tracts replicated as a function of number of repeated elements.

$(CA)_5$, $(TA)_6$, 8C stop and $(CA)_{14}$ plasmids is approximately linear with an intercept near five repeat units (Fig. 4). These data reflect experimental mutation frequencies varying from 0.4 to 13.4% with no sign of saturation. A simple hypothesis which explains these results is that the mutations are due to slippage which is enhanced as the number of repeats increases past four. Regardless of the nature of the repeats, whether TA, CA or C, the frequency of errors is proportional to their number, i.e. the 8C sequence behaves as eight repeats and $(TA)_6$ behaves as six. In these calculations we assume that the *dnaQ-G* mutation has completely eliminated exonucleolytic proofreading. Since cells with a complete loss of proofreading activity are probably inviable (26), our values for the contribution of proofreading are minimal estimates. In addition, we cannot exclude the possibility that mutated exonuclease subunits interact with the polymerase to produce holoenzyme with altered processivity or stability.

About 40–50% of the frameshift mutations made by polymerase replicating the $(CA)_{14}$ or $(CA)_5$ plasmid are corrected by wild-type proofreading exonuclease, as compared with ~80% of the errors in the 8C stop or $(TA)_6$ plasmids (Fig. 4). Based on comparisons of *dnaQmutL* and *mutL* strains, 97.5–99.5% of base substitutions are corrected by proofreading and not by mismatch repair (22). The data for the $(CA)_2$ plasmid resemble those for chromosomal base substitutions (99% corrected by proofreading). This distinction is reinforced by the different nature of the $(CA)_2$ revertants obtained (Fig. 2 and Table 4). We suppose that these +1 reversions occur at the DNA growing point and are monitored by the wild-type proofreading exonuclease in a manner similar to base substitutions. Frameshifts in highly repeated tracts are likely to occur or migrate upstream of the growing point, where they are only susceptible to mismatch repair (2). The distance from the growing point at which this slippage occurs may be related to the distance between residues in the polymerase 'thumb' (28), which fix the newly synthesized strand and its template to the protein and the end of the DNA recognized by the proofreading subunit. The longer the repeated tract, the greater the probability of a frameshift and the lower the probability that this perturbation will reach a point susceptible to nuclease action before the frameshift is irreversibly fixed by the progression of DNA synthesis. The decreased contribution of proofreading to the maintenance of long repeats *in vivo* corresponds well with the *in vitro* measurements showing decreased exonuclease surveillance of longer mononucleotide runs (29). It is possible that the major quantitative role of the mismatch repair system, at least in eukaryotes in which the DNA contains numerous dinucleotide repeats, is not the correction of base substitutions but the maintainance of the integrity of repeated regions and protection from recombination (30,31).

The statistical analyses above suggest that exonuclease deficiency has more of an effect on the 8C stop and $(TA)_6$ plasmids combined than on the $(CA)_{14}$ and $(CA)_5$ plasmids ($P = 0.02$). This difference prompts us to speculate that mononucleotides are edited more easily than dinucleotides. In addition, two different but related mechanisms may make perturbations in the $(TA)_6$ and 8C stop structure susceptible to nuclease. First, we suppose that the extrahelical nucleotide(s) can migrate towards and away from the growing point in a wave-like movement and that a mononucleotide, as in the 8C stop sequence, is likely to migrate more readily than a dinucleotide. Secondly, a TA dinucleotide is likely to melt more readily than a CA dinucleotide, making its migration towards the primer–template end easier and subjecting it to exonuclease editing. The data of Brenowitz *et al.* (32) suggest that proofreading requires melting of the newly synthesized DNA.

The stability of long repeated sequences and of single nucleotides within genes are controlled by different but overlapping systems. Proofreading enzymes may correct large numbers of polymerase-produced point mutations but be relatively inefficient at correcting errors in long repeats. Mismatch repair proteins correct errors in long repeated units (microsatellites) but in the presence of an efficient prooofreading system cells deficient in mismatch repair may remain relatively free of point mutations. It may be that the specificity of mismatch repair mutations in certain types of cancer (33,34) is related to the production of specific mutations in genes with long repeated tracts rather than to their overall effect on point mutations (35).

ACKNOWLEDGEMENTS

These experiments were started at the laboratory of Dr Tomas Lindahl at the ICRF in London and one of us (B.S.) would like to thank Dr Lindahl and the group at ICRF, particularly Barbara Sedgwick, Steve West, Rick Wood, Masahiko Sato and Bob Lloyd, for their help. We are especially grateful to Dr Barbara Bachman of the *E.coli* Genetic Stock Center and to Dr Jeffrey Miller for the bacterial strains used to start this work and to Drs Alain Sarasin and Anne Stary for the P205-GTI plasmid. We thank Dr Charles McHenry for plasmid pOPP-E and Roel Schaaper for the *dnaQ*-containing plasmid and for helpful suggestions. Dr Theodore Karrison (Department of Medicine, The University of Chicago) provided the statistical analyses and Dr Spencer Benson suggested the use of PCR for the measurement of plasmid and chromosomal DNA. Drs Roel Schaaper and Malcolm Winkler provided helpful comments on the manuscript. This work was supported in part by Department of Energy grant DE-FG02-88ER60678 and by National Cancer Institute grant R37-CA32436 and by a Fogarty International Fellowship to one of us (B.S.).

REFERENCES

1 Levinson,G. and Gutman,G. (1987) *Nucleic Acids Res.*, **15**, 5323–5338.
2 Strand,M., Prolla,T., Liskay,R.M. and Petes,T. (1993) *Nature*, **365**, 274–276.
3 Bates,H., Randall,S.K., Rayssiguier,C., Bridges,B.A., Goodman,M.F. and Radman,M. (1989) *J. Bacteriol.*, **171**, 2480–2484.
4 Sagher,D., Turkington,E., Acharya,S. and Strauss,B. (1994) *J. Mol. Biol.*, **240**, 226–242.
5 Bebenek,K., Joyce,C.M., Fitzgerald,M.P. and Kunkel,T. (1990) *J. Biol. Chem.*, **265**, 13878–13887.
6 Miller,J. (1992) *A Short Course in Bacterial Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
7 Stary,A., Menck,C. and Sarasin,A. (1992) *Mutat. Res.*, **272**, 101–110.
8 Henderson,S. and Petes,T.D. (1992) *Mol. Cell. Biol.*, **12**, 2749–2757.
9 Kunkel,T.A., Roberts,J. and Zakour,R. (1987) *Methods Enzymol.*, **154**, 367–382.
10 Hong,J. and Ames,B. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 3158–3162.
11 Nghiem,Y., Cabrera,M., Cupples,C. and Miller,J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2709–2713.
12 Drake,J.W. (1991) *Proc. Natl. Acad. Sci.USA*, **88**, 7160–7164.
13 Snedecor,G.W. and Cochran,W.G. (1980) *Statistical Methods, 7th Ed.* Iowa State University Press, Ames, IA, Chapter 15, section 13.
14 Cupples,C.G. and Miller,J.H. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5345–5349.
15 Takano,K., Nakabeppu,Y., Maki,H., Horiuchi,T. and Sekiguchi,M. (1986) *Mol. Gen. Genet.*, **205**, 9–13.
16 Blanco,L., Bernad,A. and Salas,M. (1992) *Gene*, **112**, 139–144.
17 Horiuchi,T., Maki,H., Maruyama,M. and Sekiguchi,M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3770–3774.
18 Tomasiewicz,H.G. and McHenry,C.S. (1987) *J. Bacteriol.*, **169**, 5735–5744.
19 Tomasiewicz,H.G. (1990) The Macromolecular Synthesis II Operon of *Escherichia coli*. University of Colorado Medical School, Denver, CO.
20 Schaaper,R.M. and Cornacchio,R. (1992) *J. Bacteriol.*, **174**, 1974–1982.
21 Schaaper,R.M. and Radman,M. (1989) *EMBO J.*, **8**, 3511–3516.
22 Schaaper,R.M. (1993) *J. Biol. Chem.*, **268**, 23762–23765.
23 Marians,K.J. (1992) *Annu. Rev. Biochem.*, **61**, 673–719.
24 Kang,S., Jaworski,A., Ohshima,K. and Wells,R.D. (1995) *Nature Genet.*, **10**, 213–218.
25 Balbás,P., Soberón,X., Merino,E., Zurita,M., Lomeli,H., Valle,F., Flores,N. and Bolivar,F. (1986) *Gene*, **50**, 3–40.
26 Fijalkowska,I.J. and Schaaper,R.M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 2856–2861.
27 Bebenek,K., Beard,W.A., Casas-Finet,J., Kim,H., Darden,T.A., Wilson,S.H. and Kunkel,T.A. (1995) *J. Biol. Chem.*, **270**, 19516–19523.
28 Joyce,C.M. and Steitz,T.A. (1995) *J. Bacteriol.*, **177**, 6321–6329.
29 Kroutil,L.C., Register,K., Bebenek,K. and Kunkel,T.A. (1996) *Biochemistry*, **35**, 1046–1053.
30 Matic,I., Rayssiguier,C. and Radman,M. (1995) *Cell*, **80**, 507–515.
31 Radman,M., Matic,I., Halliday,J.A. and Taddei,F. (1995) *Phil. Trans. R. Soc. Lond. B*, **347**, 97–103.
32 Brenowitz,S., Kwack,S., Goodman,M.F., O'Donnell,M. and Echols,H. (1991) *J. Biol. Chem.*, **266**, 7888–7892.
33 Parsons,R., Li,G.M., Longley,M.J., Fang,W.H., Papadopoulos,N., Jen,J., de la Chapelle,A., Kinzler,K.W., Vogelstein,B. and Modrich,P. (1993) *Cell*, **75**, 1227–1236.
34 Fishel,R., Lescoe,M., Rao,M., Copeland,N., Jenkins,N., Garber,J., Kane,M. and Kolodner,R. (1993) *Cell*, **75**, 1027–1038.
35 Markowitz,S., Wang,J., Myeroff,L., Parsons,R., Sun,L., Lutterbaugh,J., Fan,R.S., Zborowska,E., Kinzler,K.W., Vogelstein,B., *et al.* (1995) *Science*, **268**, 1336–1338.

See supplementary material available in NAR Online.